DOCUMENT RESUME

ED 281 154 CS 008 763

AUTHOR Frederiksen, Carl H.
TITLE Knowledge Representation and Discourse: Producing and Communicating Meaning.
PUB DATE 20 Mar 87
NOTE 34p.; Paper presented at the Annual Meeting of the Conference on College Composition and Communication (38th, Atlanta, GA, March 19-21, 1987).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cognitive Processes; Cognitive Structures; Computer Oriented Programs; *Discourse Analysis; Epistemology; Higher Education; Knowledge Level; *Learning Theories; Models; Prior Learning; *Reader Text Relationship; *Reading Comprehension; Schemata (Cognition); Semantics; Theory Practice Relationship; Writing Instruction; *Writing Processes

ABSTRACT

Noting that conceptions of the cognitive representations and processes involved in discourse comprehension and production recently have undergone an important shift in orientation, this paper points out the importance of the shift to the conception of the nature of texts, of text understanding, and of cognitive processes in knowledge creation and written communication. The paper indicates that the shift replaces an earlier "text-centered" view in which the linguistic structure of a text is viewed as central to discourse processing with a "knowledge-centered" view, in which semantic structures for representing knowledge and cognitive processes for generating, manipulating, and communicating knowledge are seen as central. The paper points out that the shift re-introduces an earlier perspective on text comprehension as a constructive process, but also introduces precise models of semantic representation and processing and related models for the semantic analysis of discourse. These models lead to a more unified account of discourse comprehension, production, and other discourse processing abilities, and suggest certain new instructional approaches. It then discusses (1) the shift to a knowledge-centered theory of text comprehension and production, (2) semantic grammars and discourse analysis, (3) the production of semantic representations in writing, and (4) the creation of an environment for planning and analyzing semantic structures in writing. Appendixes contain examples of semantic discourse analysis procedures and text processing schematics. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED281154

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

KNOWLEDGE REPRESENTATION AND DISCOURSE: PRODUCING AND COMMUNICATING MEANING

Paper presented at the Conference on College Composition and Communication,
National Council of Teachers of English,
March 20, 1987, Atlanta, Georgia

Carl H. Frederiksen

Laboratory of Applied Cognitive Science
McGill Cognitive Science Centre
McGill University
3700 McTavish St.
Montréal, Québec H3A 1Y2 Canada

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Carl H. Frederiksen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Conceptions of the cognitive representations and processes involved in discourse comprehension and production have been undergoing an important shift in orientation, a shift that is important to our conception of the nature of texts, of text understanding, and of cognitive processes in knowledge creation and written communication. This shift in orientation replaces an earlier text-centered view in which the linguistic structure of a text was viewed as central to discourse processing, with a knowledge-centered view in which semantic structures for representing knowledge and cognitive processes for generating, manipulating and communicating knowledge are viewed as central.

In this orientation, linguistic and textual structures are subordinated to conceptual knowledge structures in discourse processing. A text is viewed both: a) as a means of representing conceptual knowledge, and b) as a structure that incorporates strategies (implicitly or intentionally) for communicating that knowledge through its effects on the comprehension processes of its readers (or listeners).

This shift re-introduces an earlier perspective on text comprehension as a constructive process; but it introduces precise models of semantic representation and processing, and related

CS008763

methods for the semantic analysis of discourse. These models lead to a more unified account of discourse comprehension, production and other discourse processing abilities (such as translation), and they suggest certain new instructional approaches.

1. The Shift to a Knowledge-Centered Theory of Text Comprehension and Production

When cognitive research on text comprehension began in the early 1970's, attention was focused on the propositional content of text, and the inferential processes that a reader or listener applies to a text to generate propositions from text segments. Text comprehension was viewed as a process by which a reader (or listener) generates sequences of propositions, i.e., abstract semantic representations, to represent the meaning of a text in memory. These propositions are not limited to sentence meanings but included inferred propositions as well. Such *text-based* theories of comprehension viewed comprehension as a) interpreting of the literal semantic content of sentences; and b) making inferences that connect, complete, summarize, re-interpret or elaborate the literal propositional content of a text. An important component of the theories developed at this time, therefore, was the specification of the nature of propositional representations. A second component was an attempt to specify types of inferential operations or "macro-rules" that could be applied to a "propositional text base" to generate inferred propositions that summarize or embody the so-called "macro-structure" of the text.

However, it soon became apparant that text-based theories could not solve the problem of explaining why readers' text recalls contain particular inferences and not others. It was as if a reader's inferences were being directed by some top-down strategy that specified what structures the inferences were supposed to generate. To explain such selective inference, a *knowledge-based* theory was proposed in which text comprehension was viewed as a kind of pattern-recognition in which text propositions are "fit" to pre-existing patterns in memory called "frames" or "schemata". Inferences are produced in attempting to fit text information to memory frames or schemata. These are specific structures that may be described in terms of semantic network data structures

(i.e., conceptual graphs) already existing in memory that represent particular knowledge structures.

Thus, semantic representations came to be thought of as existing at two levels:

• a propositional level representing chunks of semantic information typically represented in working memory (such as events, states, or more complex relations among propositions), having modality and truth-value information, and encoded in sentences; and

• a frame level representing semantic network structures in long-term memory composed of nodes and relations linking these nodes into meaningful structural networks. According to schema theory, readers match incoming semantic information to prior knowledge structures that they retrieve from pre-existing frames residing in long-term memory. In the process they may modify these frames as necessary to fit the incoming "text data".

One problem with such "frame matching" theories is that they require specific prior knowledge of text content. If readers lack specific prior knowledge relevant to the semantic content of a text, they would have no prior knowledge frames to which to fit text propositions. Therefore they would have to rely on general-purpose text-based processing strategies. However, research on comprehension of unfamiliar stories and expository texts suggests that this is not the case. Evidence of selective inferential processing in recall and on-line text interpretation tasks indicates that readers use top-down strategies for generating frame representations of a text even in the absence of relevant prior knowledge. Furthermore, they have specialized competency in understanding particular types of semantic networks (i.e. frame structures) expressed in texts (such ad narrtives, procedures, problems, etc.).

To account for this ability to generate semantic representations, I have been developing a second type of knowledge-based theory of text processing that may be termed a *rule-based* theory. In a rule-based theory, a comprehender is viewed as possessing rules for forming particular types of semantic structures expressed in discourse. These rules are assumed to be characterizable as semantic grammars A reader is

assumed to "semantically parse" a text at different levels by applying the production rules of the grammar to generate semantic structure. This occurs both at a propositional level, and at higher levels in which frame representations are constructed.

In current rule-based theory, semantic grammars are not attached to any linguistic or sociolinguistic category (such as was the case with "story grammars"; rather, they are sets of language-independent rules for defining a type of semantic representation. The rules in our grammars are expressed in the form of "BNF grammars", a formalism taken from computer science that corresponds to a recursive transition network parser. Consequently, the grammars have semantic parsers associated with them which may be proposed as models of human text processing.

2. Semantic Grammars and Discourse Analysis

Semantic grammars have been developed for both propositions and for different types of frame-level semantic structures. Propositions represent discrete chunks of semantic information that are encoded syntactically and lexically in language. Frames are semantic structures that reflect different types of semantic networks that are used to represent declarative information in long-term memory (such as the structure of narratives, procedures, descriptions, or problems). Semantic networks are structures for data representation that are composed of nodes (that usually contain concepts or pointers to other semantic network structures) and links (labelled arcs or arrows pointing from one node to another). Propositions, which are usually represented by a predicate + arguments formalism, are equivalent to semantic networks.

Associated with these grammars are semantic discourse analysis procedures in which the grammars are applied to analyze a text , first at a propositional level, and second, at a frame level. This will be illustrated presently using scientific writing in the domain of chemistry.

The semantic grammars being used in discourse analysis are currently being implemented on Macintosh microcomputers as "electronic grammars" which enable a researcher to analyze

discourse while guided by a computer through a "top-down" semantic parse of text segments. Using computer discourse anlaysis environments to implement grammars, it is possible to create tools that enable the precise semantic analysis of text according to rules specified in a semantic grammar (e.g., a propositional grammar). Previous methods of "propositional analysis" were necessarily approximate and intuitive since the rules underlying the representations were not specified precisely. We are constructing a graphic environment for conducting such analyses. The idea is to create an instructional environment for training and assisting writers in planning and analyzing semantic structures in composing and revising texts. Semantic structures are displayed as objects on the screen that can be manipulated.

Semantic discourse analysis procedures may be illustrated using a text on the chemistry of photosynthesis (Figure 1).

[Insert Figure 1 about here]

This text was segmented into clausal units consisting of main clauses and bound adjuncts. Propositions were generated for each clausal unit by applying the propositional grammar to the segment. Any propositional information that was syntactically or lexically marked in a segment was introduced into the proposition (Figure 2).

[Insert Figure 2 about here]

Then, descriptive frames were generated from the propositional representation of the (Figures 3-6). In so doing, local coherence inferences involving slot filling and replacement of anaphoric elements with their referents was first carried out.

[Insert Figures 3-6 about here]

The resulting descriptive frames provide an abstract representation of knowledge of a topic in chemistry. This "expert frame" may be used as a canonical representation to assess students' knowledge of this topic in chemistry, and their ability to use this knowledge (e.g., in interpreting experiments), and communicate it. The frame representation is abstract in the sense that it is only indirectly related to the text. The specification of how a text can be related to a semantic frame structure provides a

starting point for analyzing the processes involved in the production of semantic and discourse structures in writing. In the present account, I will emphasize the production of semantic representations. Bracewell has discussed the linkage between text generation and the generation of semantic structure in composing discourse.

3. The Production of Semantic Representations in Writing

Despite the fact that rule-based models for semantic representation were developed to study text comprehension, they have shifted the orientation from text processing to knowledge generation. This orientation to text processing is depicted in the diagram (Figure 7).

[Insert Figure 7 about here]

Here, different levels of semantic ("conceptual") and language ("textual") representation are indicated together with arrows indicating how one level of structure is derived from other levels.

In discourse comprehension ("reversing" the arrows for a moment), a reader is confronted with a text consisting of a linear sequence of sentences containing cohesive elements and topicalization patterns that define a textual surface structure. The reader must be able to generate propositions for each text segment (sentence or clause), carry out "local coherence inferences" to connect propositions to earlier ones in the sequence and fill missing information in slots, and finally generate a non-linear semantic network structure (i.e., a frame) as a reconstruction of the "semantic model" underlying the text. The ease with which a non-linear frame structure can be generated will depend on how frame information was sequenced, the explicitness of the propositional representation of frame information, the quality and complexity of the encoding of propositions in clause structures, and the manner in which topicalization, cohesive relations and other features of text surface structure relate to the frame structure .

In discourse production, this process is reversed: the writer begins by retrieving and/or generating a frame representation and must make a series of decisions that eventually result in a textual

representation of the frame. This "mapping" is one-to-many (the writer is continuously faced with choosing among many alternative representations). Furthermore, text production occurs in real time: the writer must continually generate semantic and language representations that satisfy multiple constraints. No account of the writing process can be complete that does not account for how writers' develop knowledge production and communicative strategies to control these choices.

To illustrate processes involved in the generation of semantic structures in writing, I will use sample text productions from a current study in which high school students familiar with a word processing program (the Bank Street Writer) were asked to write procedural texts providing instructions to other students not yet familiar with this word processor. To develop a canonical model of the procedural frame for using the Bank Street Writer, we applied frame analysis procedures to a text used in the school that summarizes procedures for using the Bank Street Writer (Figures 8-11).

[Insert Figures 8-II about here]

An example of a student writer's procedural text is given in Figure 12.

[Insert Figure 12 about here]

This student's text was analyzed by identifying information from the procedural frame that was identified. In the next Figures, the procedural nodes and links represented in the student's text are indicatedin the sequence in which they were produced (Figures 13-16).

[Insert Figures 13-16 about here]

Thus, the manner in which the writer selected frame information, and the linear sequence used to communicate it are reflected in the Figure. This student identified connected frame structures and established a coherent strategy for sequencing this information in text. Further propositional analysis of this student's text would reflect the student's generation of propositions, and subsequent analysis of clause and text surface structures would reflect

strategies for encoding and signalling frame information in text.

In summary, we are interested in how expert writers coordinate their processes of frame generation, linearization, proposition generation, encoding and textual staging in on-line text generation, and eventually, in teaching novice writers to use these strategies. The prior specification of frame structures (when texts are being written in delimitable domains of knowledge) has proven to be effective in aiding the analysis of the text production process. Furthermore, by studying readers' comprehension of texts embodying different production strategies at these various levels, we can directly connect expert writers' strategies to their communicative effects on various types of readers.

4. An envirnoment for planning and analyzing semantic structures in writing.

Our observations of student writers suggest that they appear to have difficulty with the specification and manipulation of the semantic content of a text. Frequently the semantic representation (frame structure) is inadequately developed, and students seem to lack an ability to modify the content of their writing. Students do appear to be able to modify sentence structures and textual features of their writing, but not in a way that is related to the function and organization of sentences to communicate meaning.

Thus we have what appears to be a paradoxical situation in which, although one expects that the primary function of language is to communicate meaning, students appear to be unable to manipulate their meanings while they are able to manipulate sentence structures and text surface structures! One explanation of this apparant anomaly is that text units appear as objects that can therefore can be manipulated, while "ideas" or meanings are not represented directly as objects, but only as they have been encoded in language and arranged in texts. What would appear to be required, therefore, is a means for representing and manipulating semantic objects. Such an environment would provide an aid to analyzing problems with documents, and to revising and composing.

To explore this possibility, we are developing an object-oriented environment for the representation of semantic structures

graphically as objects on a computer screen. We are developing three types of instructional environments based on the representation and manipulation of semantic structures:

a) tutors for learning to carry out semantic analysis (parsing) of texts for propositions and frame-level semantic structures (The purpose of semantic tutors is to provide students with explicit tools for representing and analyzing the semantic structures of written text, and an awareness of the variety of sentence/text structures that can be used to express a given semantic structure.);

b) an environment for applying these tools to analyse and revise texts (The text analysis environment will enable students to analyze their own (or any) texts and modify their texts on the basis of their analyses. This could include modifying the semantic frame information, the means by which this is made explicit through propositional instantiation, and the linearization and communication of semantic information in sentence and text structure.);

c) an environment for developing semantic representations, planning, and composing texts from semantic representations (This "semantic workbench" production environment will enable a student to compose and manipulate graphic frame representations on the screen, explore linearization strategies, and create sentences that vary in the propositions they encode and the lexical, syntactic and textual structures they employ.

A writer needs to be able to independently manipulate different component representations in text production (as specified by the rule-based cognitive theory of production) and to understand how manipulating one level entails changes in productions at other levels. The discourse production environment will assist a student writer in developing facility in creating and manipulating semantic and discourse structures, and in understanding the relationships among these representations that are involved in rhetorical choice.

References:

Frederiksen, C.H. (1986) Cognitive models and discourse analysis. In C.R. Cooper & S. Greenbaum (Eds.), Written communication annual, Vol. I: Linguistic approaches. Beverly Hills, Sage.

Frederiksen, C.H. (1987) Constructing and manipulating multi-level semantic representations in comprehending and producing text. In G. Denhière and J. Rossi (Eds.), Texts and text processing: Psychological and linguistic aspaects. (book in preparation).

Frederiksen, C.H. (1987) Exploring cognitive processes in text processing within functional contexts. In D. Bloom (Ed.), Learning to use literacy in educational settings. Norwood, NJ: Ablex (book in preparation).

Frederiksen, C.H. (1987) Evaluating knowledge acquisition and use. In N. Frederiksen, R. Glaser, A. Lesgold & M. Shafto (Eds.), Diagnostic monotoring of skill and knowledge acquisition. Hillsdale, NJ: Erlbaum (book in preparation).

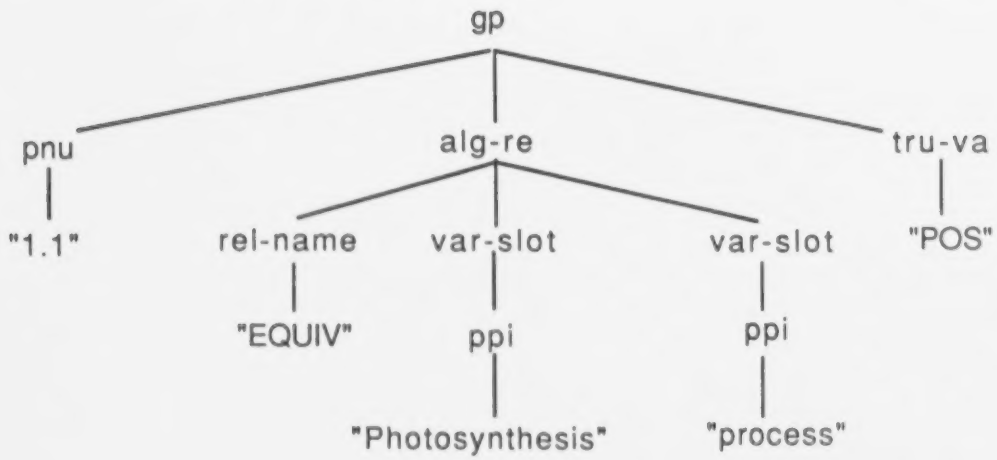
Frederiksen, C.H., Donin-Frederiksen, J. & Bracewell, R.J. (1986) Discourse analysis of children's text production. In A. Matsuhashi (Ed.), Writing in real time. Norwood, NJ: Ablex.

Bracewell, R.J. (1987) Signaling semantic representation: The role of topical patterning in text. Paper presented at the fourth Annual International Conference on the Teaching of English, Ottawa.

Input File to CODA Produced by CODAPR -
Photosynthesis Expert Text (Paragraph 1)
(photo.inp)

(Note - < > 's enclose words that appear as nodes in CODA parsing tree)

<Photosynthesis> is the <<process>>
by which <green> <<plants>> <harness>
the <energy> of <sunlight>
<absorbed> by <chlorophyll> to
<build> <organic> <compounds> from
<carbon_dioxide> and <water> .
--1--------------
The reaction is often referred
to as assimilation of carbon .
--2--------------
Photosynthesis has a
fundamental role to provide
organic food which is needed
by all living beings , plants
and animals , for their
existence , growth and
multiplication .
--3--------------
If photosynthesis did not
exist , all living beings
would disappear from the earth
within the lifetime of a
human generation .
--4--------------

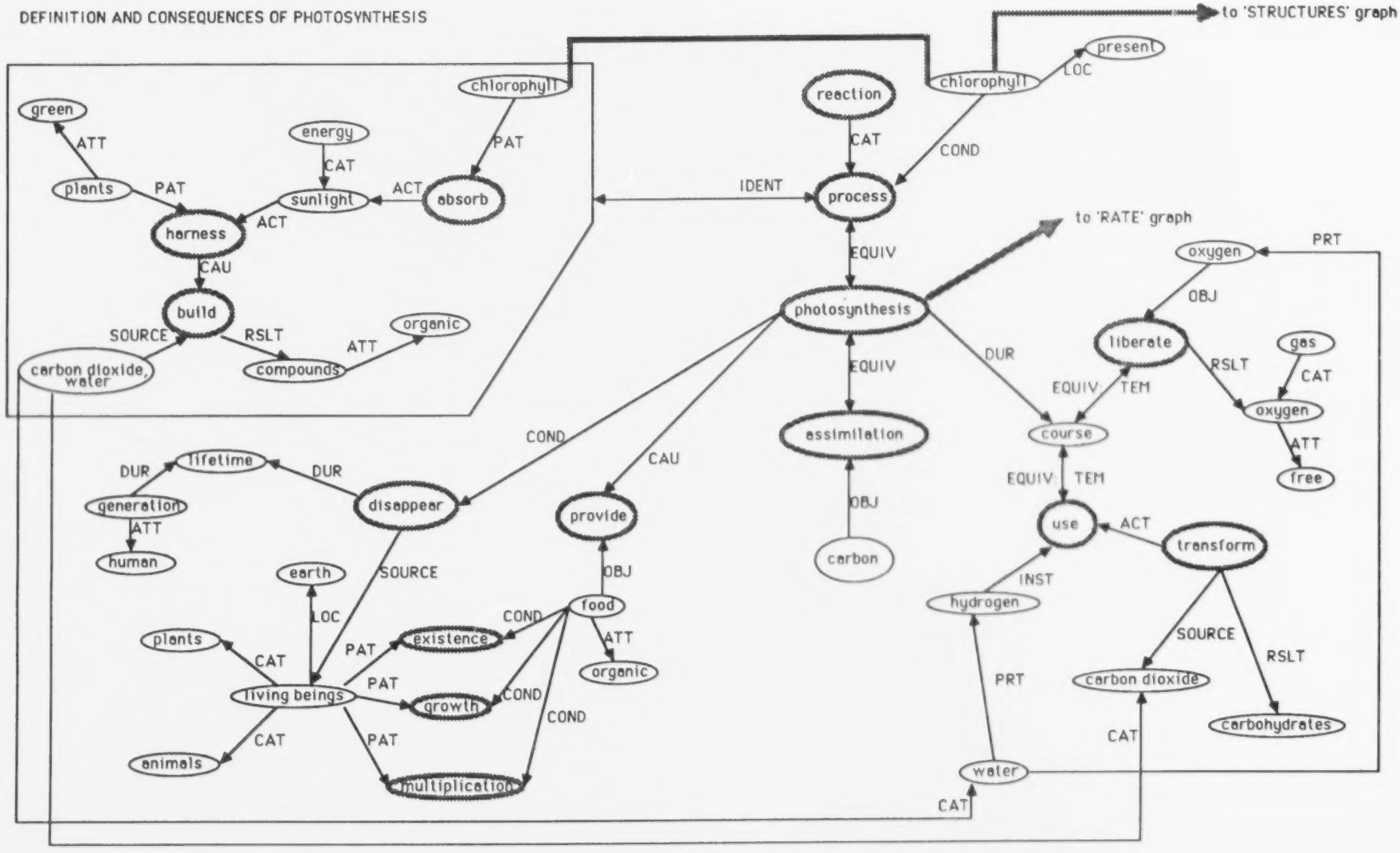


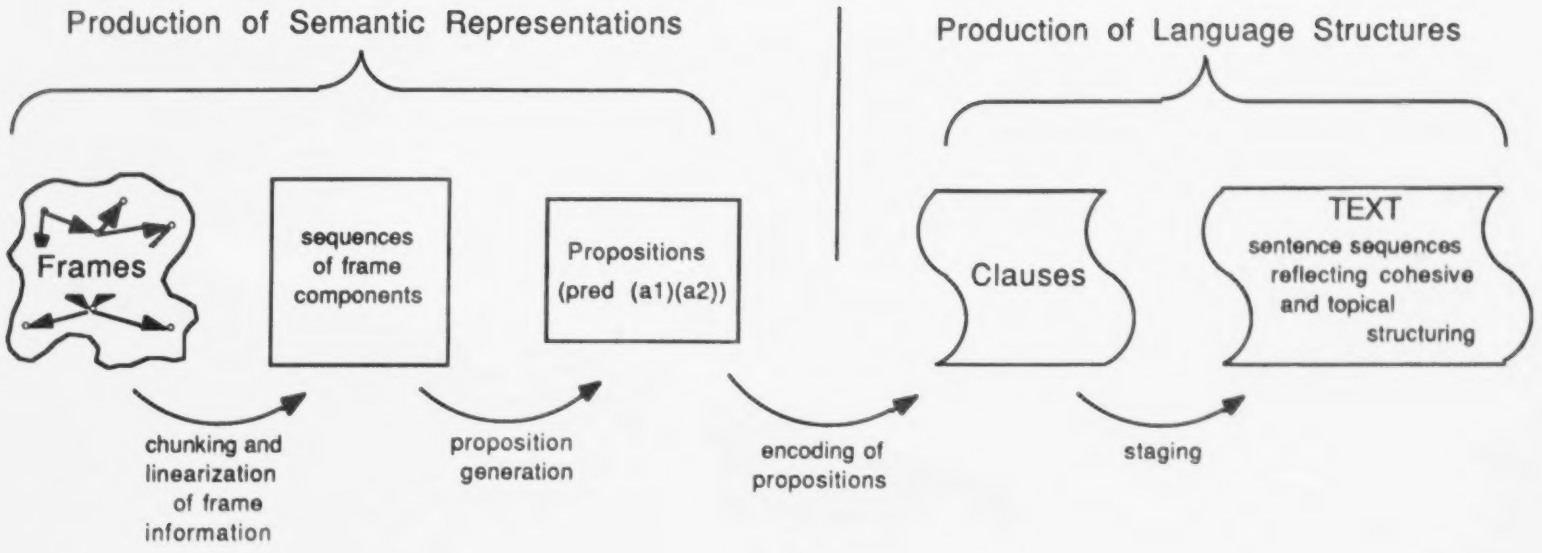
Parsing Tree for Proposition 1.1

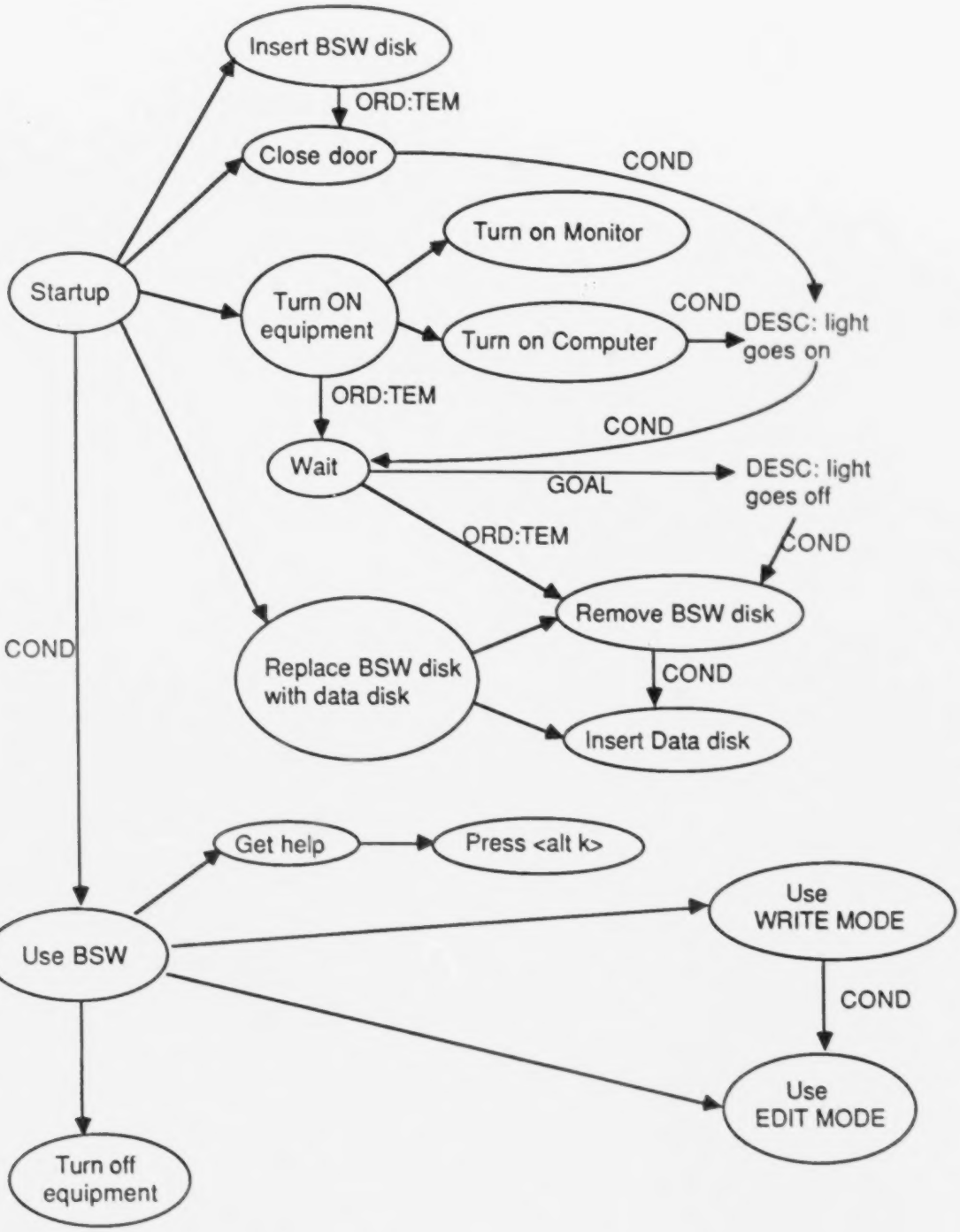
Propositional Representation: Predicate-Argument Notation
Segment 1 - Photosynthesis Expert Text

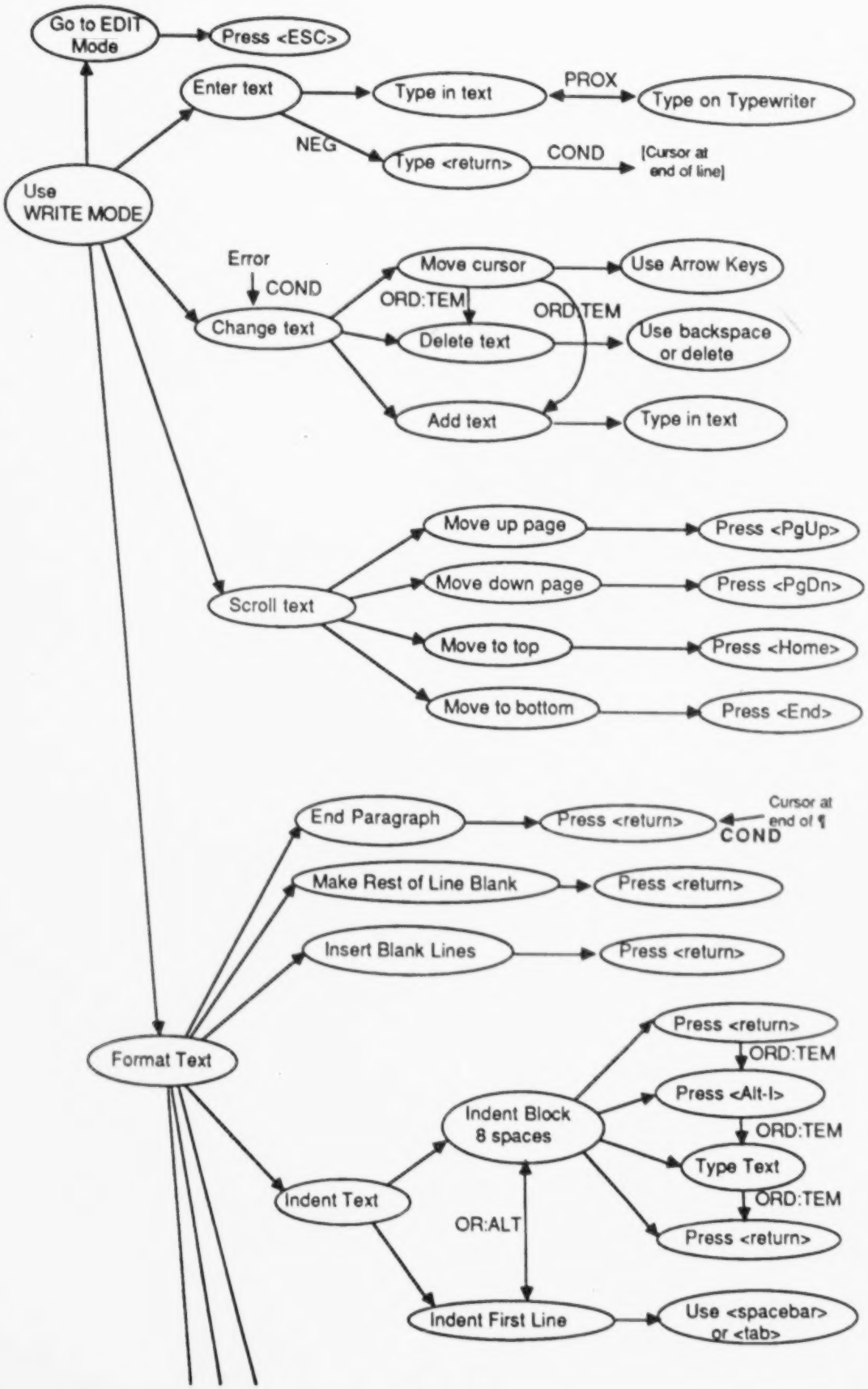
1.1	EQUIV	["Photosynthesis"], ["process"], POS;
1.2	IDENT	["process"], [1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9], POS;
1.3	"harness"	pat-re: plants, act-re: 1.4, POS;
1.4	"energy"	cat-re: "sunlight", POS;
1.5	"absorb"	pat-re: "chlorophyll", act-re: 1.4, POS;
1.6	"plants"	att-re: "green", POS;
1.7	CAU	[1.3], [1.8];
1.8	"build"	sour-re: "carbon_dioxide", "water," resu-re:1.9, POS;
1.9	"compounds"	att-re: "organic", POS;
2.1	EQUIV	["reaction"], [2.2], tem-re: "often", POS;
2.2	"assimilation"	obj-re: "carbon", POS;
3.1	CAU	["photosynthesis"], ["role"], POS;
3.2	IDENT	["role"], [3.3, 3.4], POS;
3.3	"provide"	obj-re: "food", POS;
3.4	"food"	att-re: "organic", POS;
3.5	"role"	att-re: "fundamental", POS;
3.6	"need"	pat-re: "living_beings", obj-re: "food", POS;
3.7	COND	[3.4], [3.8, 3.9, 3.10], POS;
3.8	"existence"	pat-re: "their", POS;
3.9	"growth"	pat-re: "their", POS;
3.10	"multiplication"	pat-re: "their", POS;
3.11	"living beings"	cat-re: "plant", POS;
3.12	"living beings"	cat-re: "animals", POS;
4.1	"exist"	act-re: "photosynthesis", NEG;
5.1	"disappear"	sour-re: 5.2, dur-re: "lifetime", COND, POS;
5.2	"living beings"	loc-re: "earth", POS;
5.3	"generation"	dur-re: "lifetime", POS;
5.4	"generation"	att-re: "human", POS;
5.5	COND	[4.1], [5.1, 5.2, 5.3, 5.4], POS;

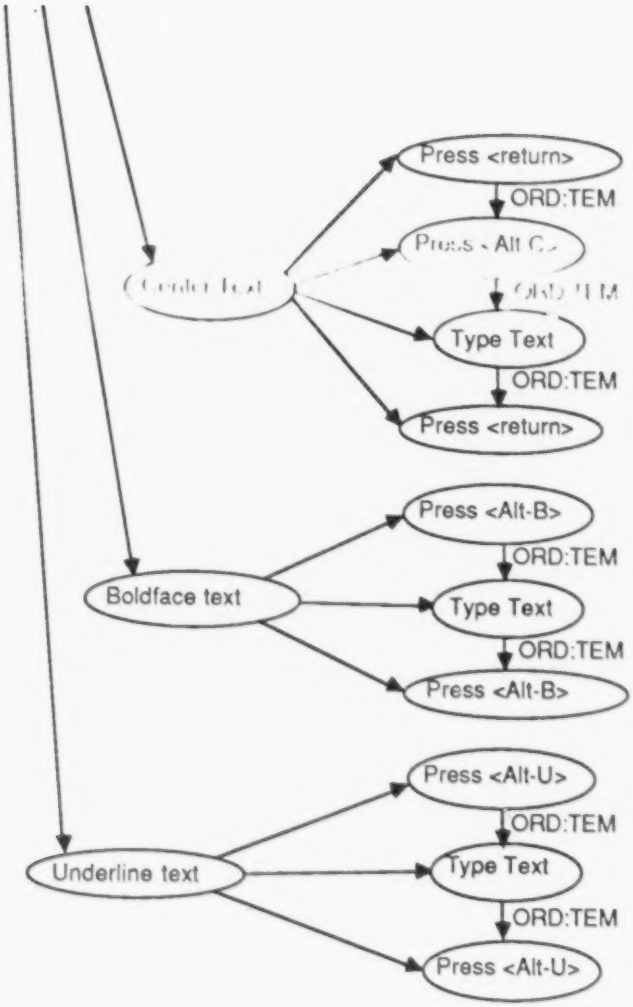
DEFINITION AND CONSEQUENCES OF PHOTOSYNTHESIS

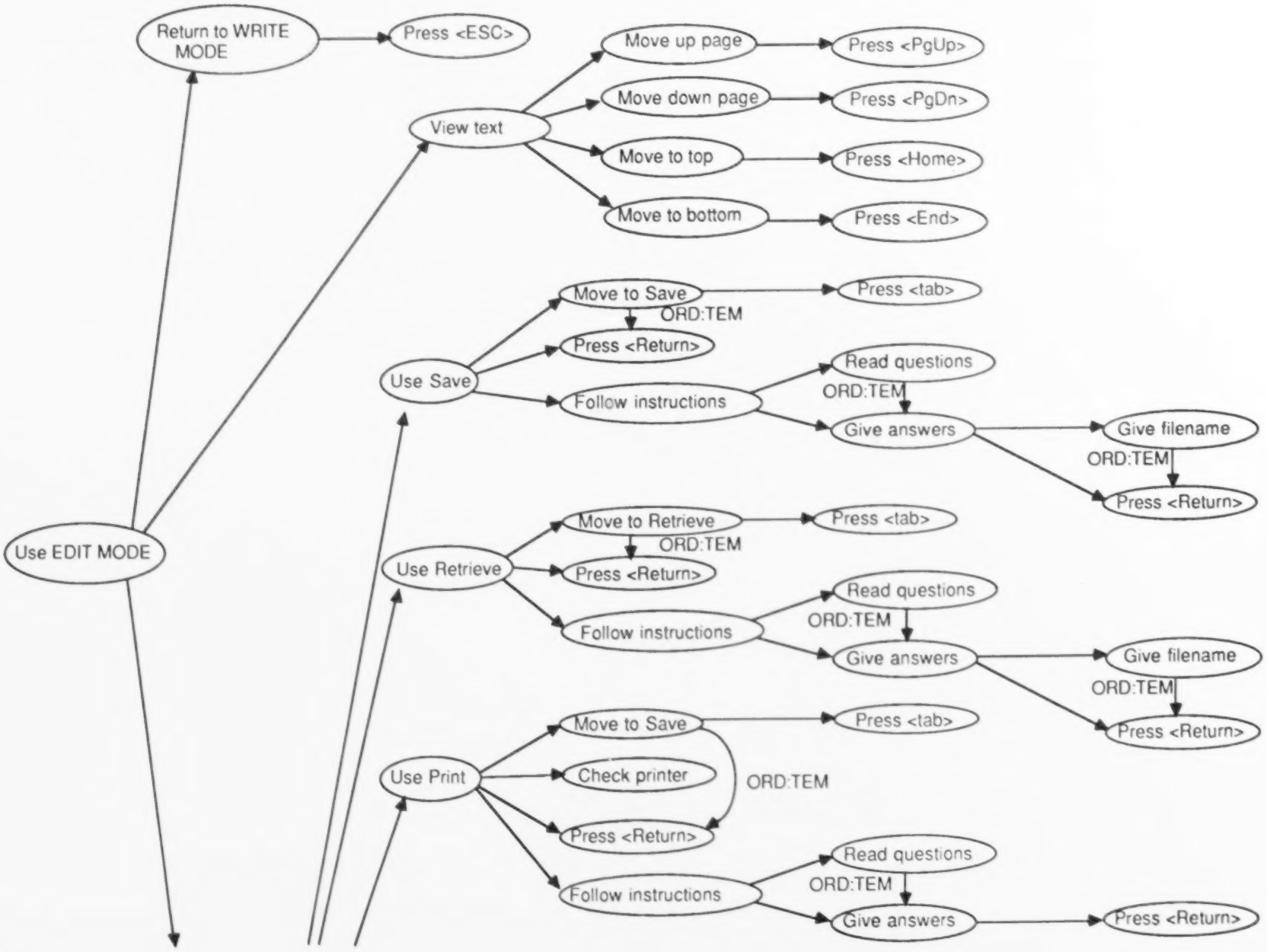


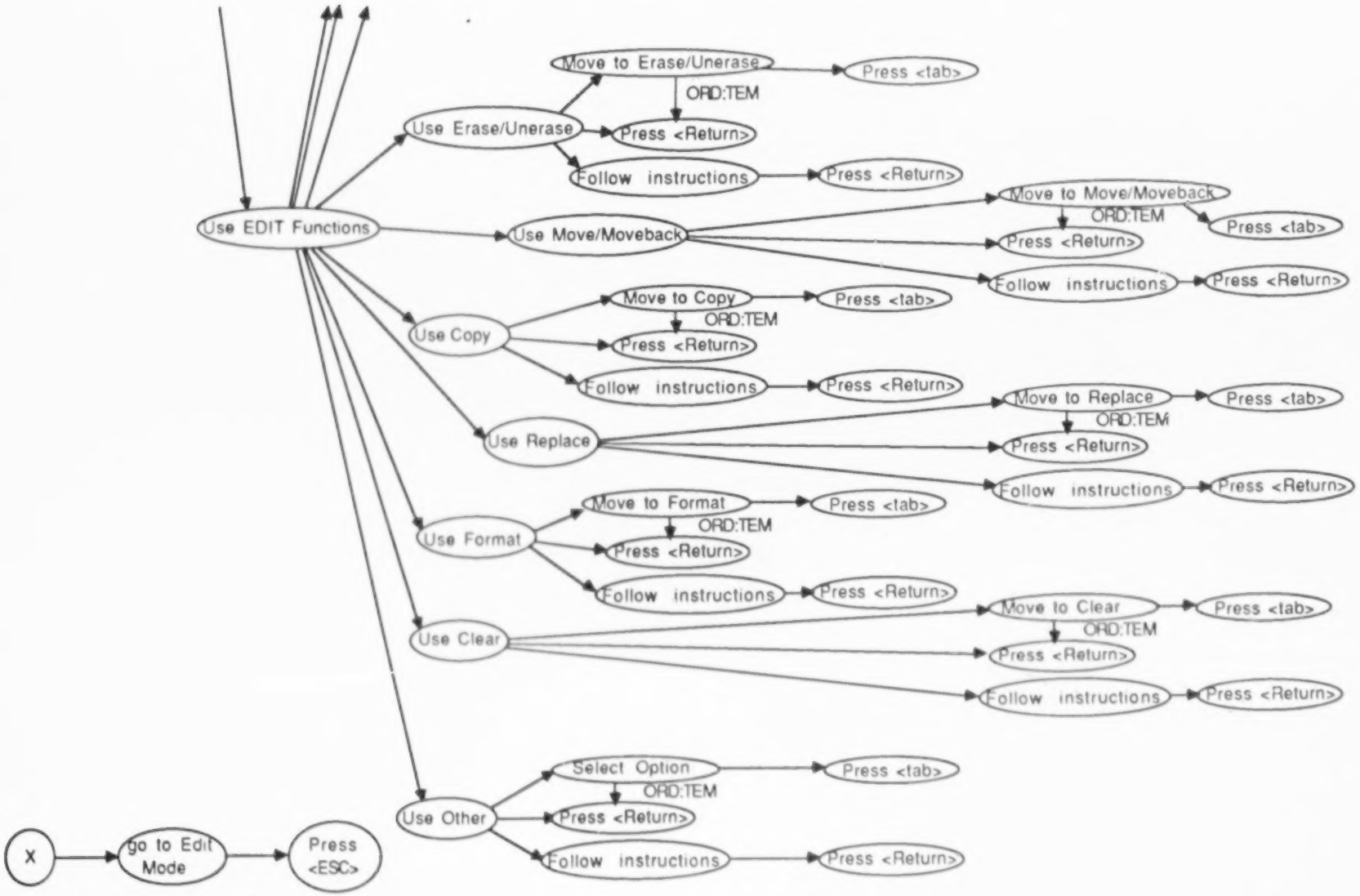

Production of Semantic Representations
Production of Language Structures
Frames
sequences of frame components
Propositions (pred (a1)(a2))
Clauses
TEXT
sentence sequences reflecting cohesive and topical structuring
chunking and linearization of frame information
proposition generation
encoding of propositions
staging


Insert BSW disk
ORD:TEM
Close door
COND
Turn on Monitor
Startup
Turn ON equipment
COND
Turn on Computer
DESC: light goes on
ORD:TEM
COND
Wait
GOAL
DESC: light goes off
ORD:TEM
COND
Remove BSW disk
COND
Replace BSW disk with data disk
Insert Data disk
COND
Get help
Press <alt k>
Use BSW
Use WRITE MODE
COND
Use EDIT MODE
Turn off equipment


Go to EDIT Mode
Press <ESC>
Enter text
Type in text
PROX
Type on Typewriter
NEG
Type <return>
COND
[Cursor at end of line]
Use WRITE MODE
Error
COND
Change text
Move cursor
Use Arrow Keys
ORD:TEM
ORD:TEM
Delete text
Use backspace or delete
Add text
Type in text
Scroll text
Move up page
Press <PgUp>
Move down page
Press <PgDn>
Move to top
Press <Home>
Move to bottom
Press <End>
End Paragraph
Press <return>
Cursor at end of ¶
COND
Make Rest of Line Blank
Press <return>
Insert Blank Lines
Press <return>
Format Text
Press <return>
ORD:TEM
Press <Alt-I>
ORD:TEM
Indent Block 8 spaces
Type Text
ORD:TEM
Indent Text
OR:ALT
Press <return>
Indent First Line
Use <spacebar> or <tab>


Press <return>
ORD:TEM
Type Text
ORD:TEM
Press <return>
Press <Alt-B>
ORD:TEM
Boldface text
Type Text
ORD:TEM
Press <Alt-B>
Press <Alt-U>
ORD:TEM
Underline text
Type Text
ORD:TEM
Press <Alt-U>


Use EDIT MODE
Return to WRITE MODE
Press <ESC>
View text
Move up page
Press <PgUp>
Move down page
Press <PgDn>
Move to top
Press <Home>
Move to bottom
Press <End>
Use Save
Move to Save
Press <tab>
ORD:TEM
Press <Return>
Follow instructions
Read questions
ORD:TEM
Give answers
Give filename
ORD:TEM
Press <Return>
Use Retrieve
Move to Retrieve
Press <tab>
ORD:TEM
Press <Return>
Follow instructions
Read questions
ORD:TEM
Give answers
Give filename
ORD:TEM
Press <Return>
Use Print
Move to Save
Press <tab>
Check printer
ORD:TEM
Press <Return>
Follow instructions
Read questions
ORD:TEM
Give answers
Press <Return>



Where X = {set of nodes in Edit mode graph not labelled with "Use"}

Table 3

Protocol of Subject RW1.2112R

Bank Street Writer

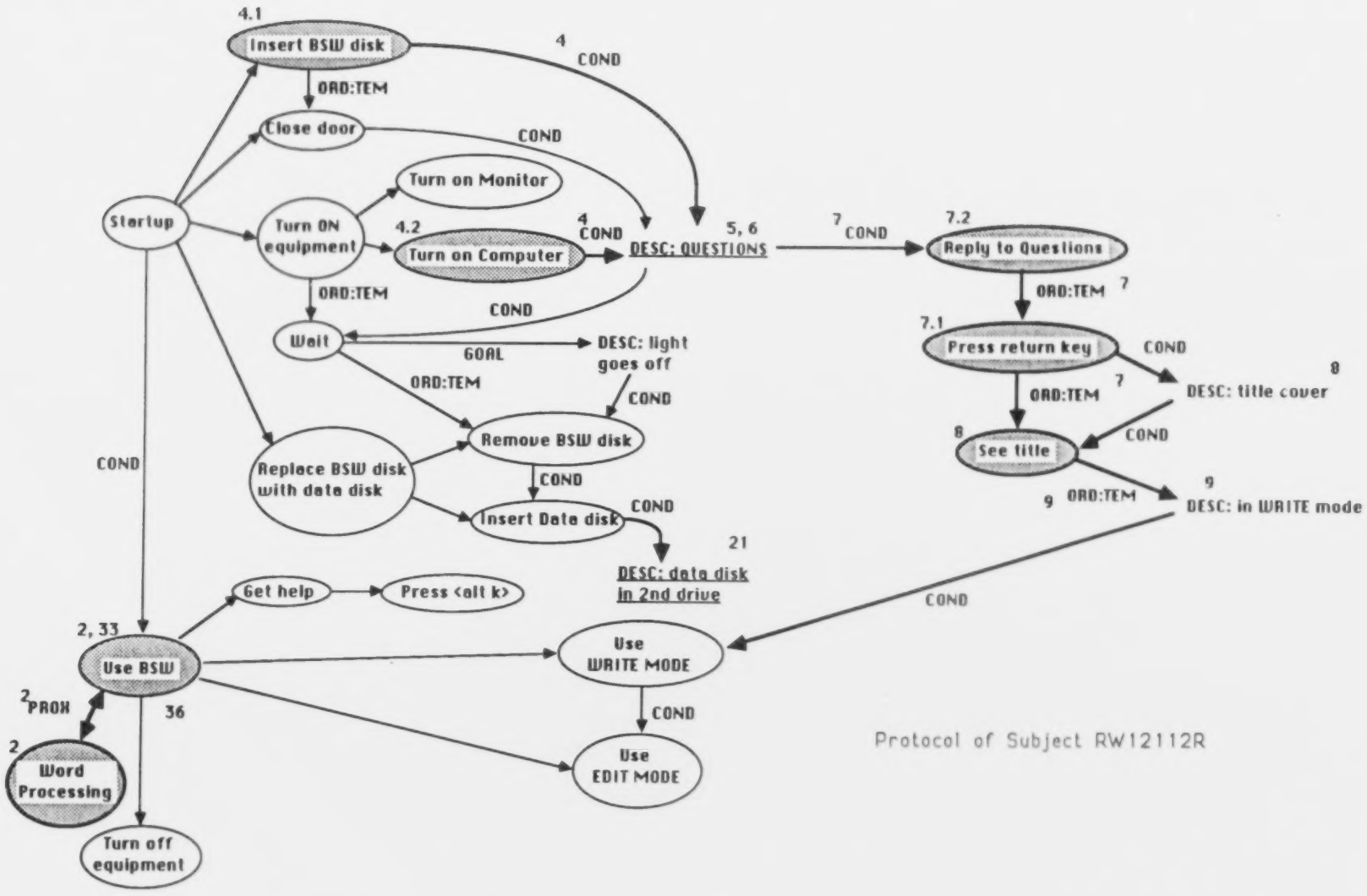
Bank Street Writer is a program that allows you to correct mistakes, add more words, delete words, erase words, move paragraphs around in a text that you have typed in, without having to type the text over. 1/
Bank Street Writer is like a word processing. 2/
B.S.W. is short for Bank Street Writer. 3 /
When you put the B.S.W. in the disk drive and turn on the computer 4 /
it will ask you for the date, starting from the month, date and year. 5 /
Then it will ask you for the time 6/
after you have pressed the return key after everyone of your replies.7 /
You will see a title cover on the screen 8/
and then you will be in the writing mode.9 /
On the top of the screen you will see instructions. 10/
Before typing in your input you should read the instructions. 11/
{After reading the instructions you can type in your text.}
After you have finished typing your text 12/
you can press ESP 13/
and that will get you to the editing functions. 14 /
You get to the function that you want by using the arrow key underneath ESP.15/
To save you highlight SAVE16/
and press return.17/
On the screen there should appear questions like what do you want the file to be called and etc.. 18 /
Answer the questions 19/
and then your input should be saved20/
if you have your disk in the second disk drive. 21 /
To print, you go back to the function mode 22/
and highlight PRINT.23/

Table 3 (cont.)

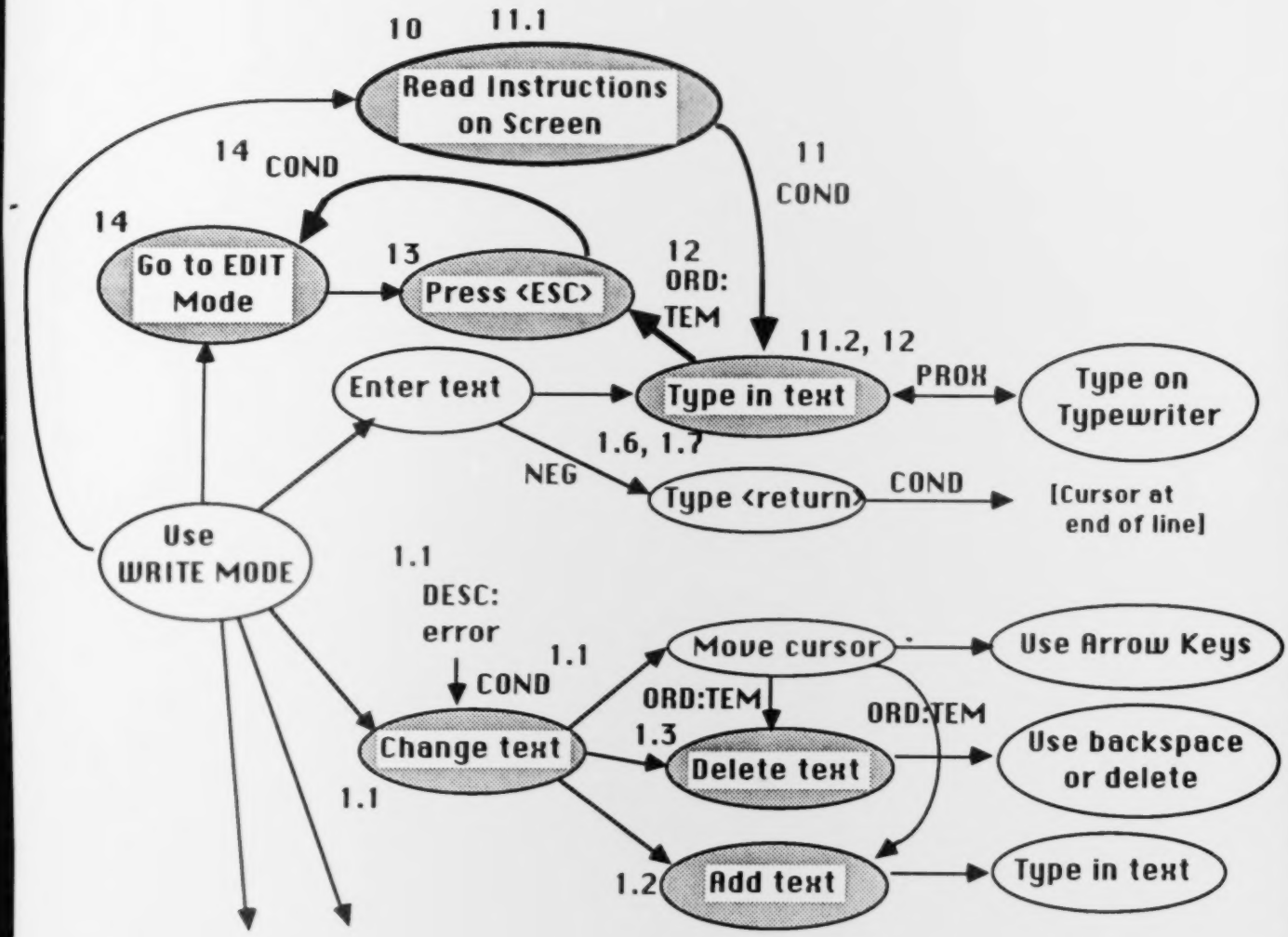
Answer the questions,[24]/
but before you should turn the computer channel to your computer,[25]/
turn on the printer,[26]/{and}
fix the paper[27]/
and then you're ready to print. [28] /
After answering all the questions the printer should start printing[29]/.

To retrieve, you highlight RETRIEVE in the function mode [30]/
and answer the questions that the machine {computer} asks. [31] /
It inquires things {asks queestions } like which input do you want to retrieve. [32]/

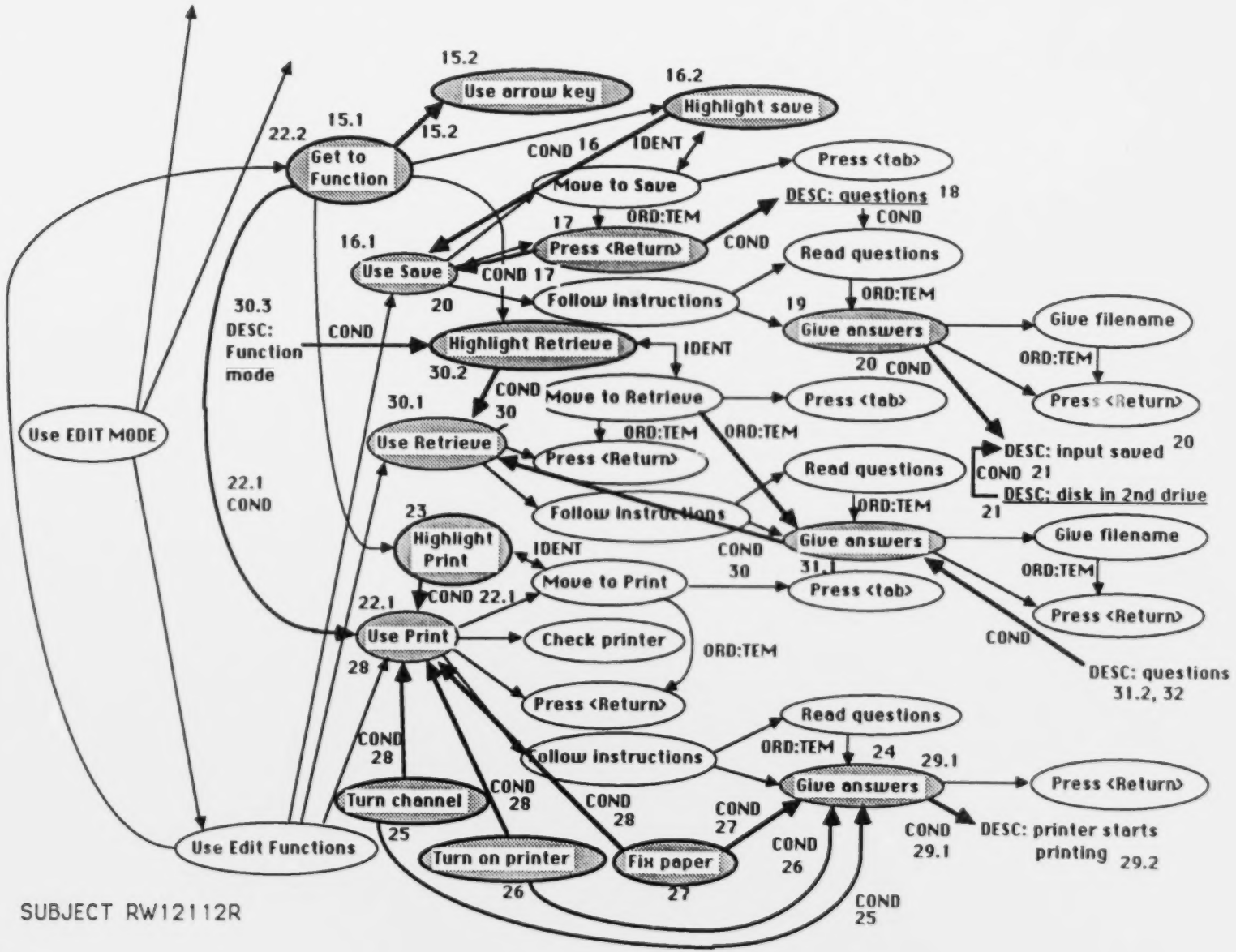
The Bank Street Writer is a program that I enjoy doing [33]/
because it isn't very complicated. [34] /
I think that with this program it saves time[35] /
and it is fun to use.[36] /



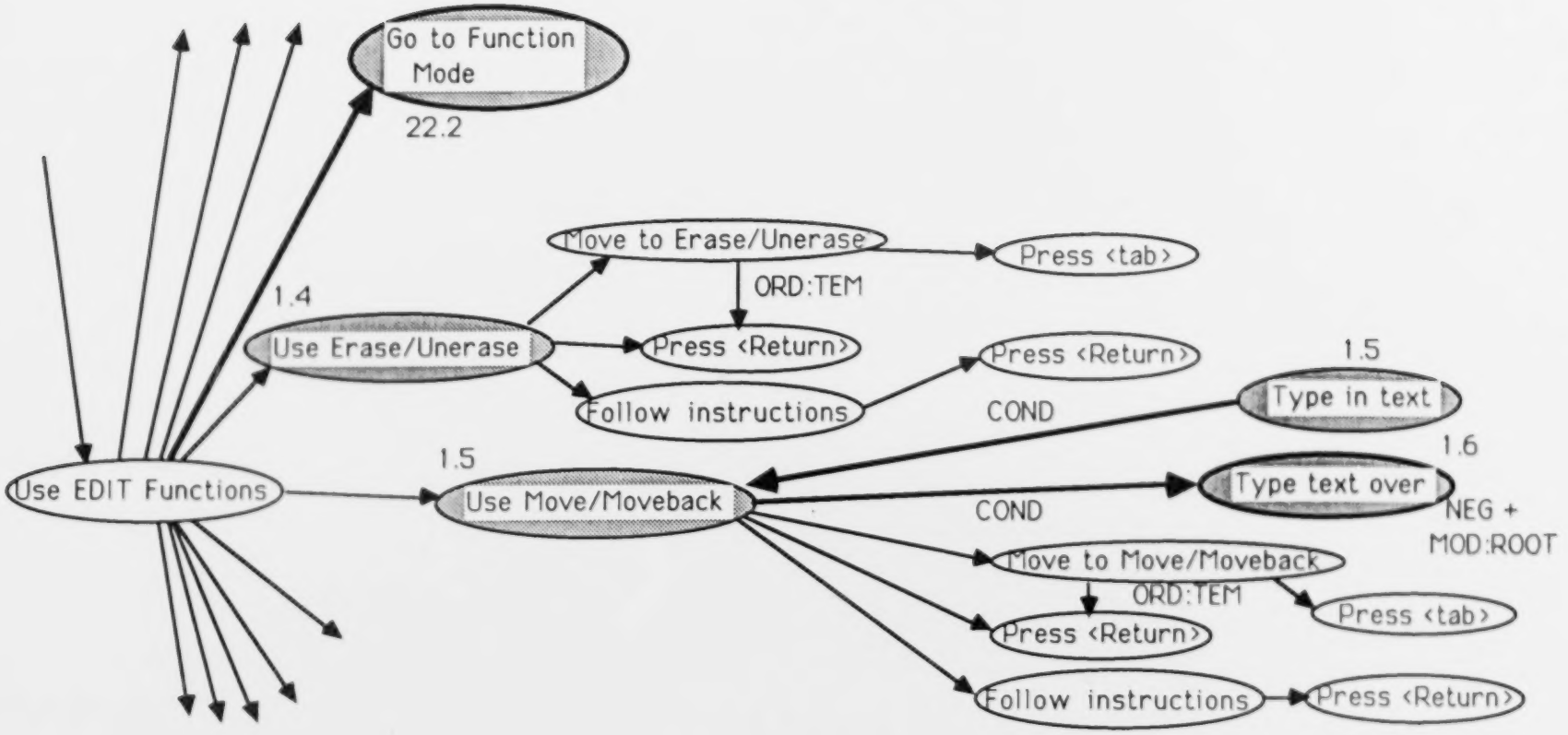
Protocol of Subject RW12112R



Protocol of Subject RW12112R



SUBJECT RW12112R



Protocol of Subject RW12112R